(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁷: **G01N 21/49**, A61B 5/00

(21) International Application Number: PCT/GB02/00508

(22) International Filing Date: 6 February 2002 (06.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0103030.3   7 February 2001 (07.02.2001)   GB

(71) Applicant: **UNIVERSITY COLLEGE LONDON** [GB/GB]; Gower Street, London WC1E 6BT (GB).

(72) Inventor: **PICKARD, David**; National Medical Laser Centre, 1st Floor, Charles Bell House, 67-73 Riding House Street, London W1P 7DN (GB).

(74) Agent: **ELKINGTON AND FIFE**; Prospect House, 8 Pembroke Road, Sevenoaks, Kent TN13 1XR (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD OF PROCESSING A BROADBAND ELASTIC SCATTERING SPECTRUM OBTAINED FROM TISSUE
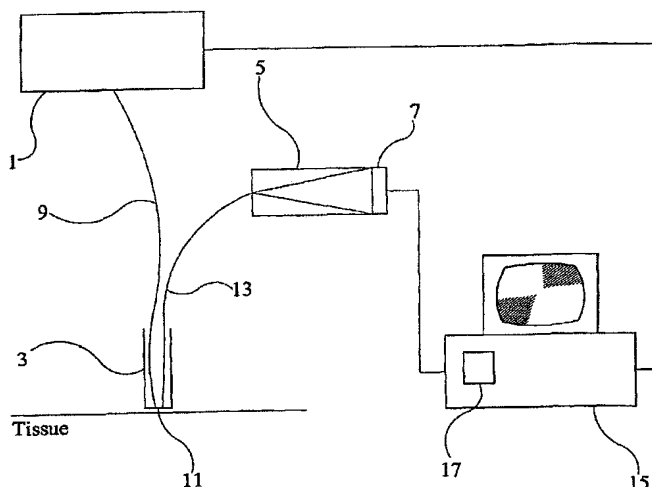
(57) **Abstract:** A method of processing a broadband optical elastic scattering spectrum obtained from tissue comprises the steps of: obtaining, in a plurality of fitting ranges of wavelength, fitting parameters giving a best fit to the spectrum in the respective fitting ranges; and recording the fitting parameters as a parameter set representing the spectrum; wherein in at least one fitting range the fit is to the absorption of at least one predetermined component, for example, haemoglobin, and in the reminder of the fitting ranges the fit is to a smooth function. The fitting parameters may be used for classifying the tissue, preferably using hierarchical cluster analysis. An apparatus for obtaining elastic scattering spectra from tissue and adapted to process the spectra using the above method is also claimed.

WO 02/063282 A1

METHOD OF PROCESSING A BROADBAND ELASTIC SCATTERING SPECTRUM OBTAINED FROM TISSUE

The invention relates to a method of processing a spectrum, in particular an elastic scattering spectrum taken from tissue and to apparatus including a spectrum processor for carrying out the method.

Elastic scattering spectroscopy is a known technique for investigating tissue. In essence, light is shone into human tissue, generally living human tissue, and a photoreceptor measures the light transmitted to the photoreceptor through scattering in the tissue. The spectrum of light passing through the tissue is then recorded, and used to assist in diagnosis of any of a number of medical conditions that the patient may have. Thus, the technique may be described as optical biopsy.

Prior art apparatus for carrying out optical biopsy is presented in WO98/27865 to David Benaron, and in US 5,303,026 to Stroble et al. The latter patent describes a system having a light source feeding into a reference optical fiber and a probe optical fiber. The probe optical fiber being brought to a probe tip. The probe tip has another optical fiber arranged adjacently of it, which collects light and brings it to a detection system which compares its intensity to the intensity of light on the reference optical fiber. When the probe tip is brought against human tissue the detection system can record the difference as a between the reference signal strength and that of the light scattered by the human tissue as a function of wavelength to obtain an optical biopsy spectrum.

The use of an elastic scattering spectrum to diagnose a number of medical conditions is described in a number of papers. Zhengfang GE et al describe in the paper "Identification of Colonic Dysplasia and Neoplasia by Diffuse Reflectance Spectroscopy and Pattern Recognition techniques", Applied Spectroscopy Volume 52 number 6 (1998) p 833, a method of identifying colonic dysplasia and neoplasia. The paper describes a number of different pattern recognition techniques used to evaluate the samples.

2

One of these is multiple linear regression analysis, which is used to fit to reflectance intensities measured at 26 different wavelengths every 16nm in the range 350nm to 750nm. An output score is obtained from the formula:

$$score = k + \sum_{j=1}^{26} a_j D_i(\lambda_j)$$

5   The coefficients a are fitted coefficients arranged such that the score is +1 for adenomatous polyps and -1 for hyperplastic polyps. $D_i\lambda_i$ is the reflectance value for the ith tissue sample at the jth wavelength.

Another approach described in the paper by Zhengfang et al is linear discriminant
10   analysis. This is a method of classifying a test into one of k groups using a classification score that can be computed from a formula. The test is classified into the group which gives the lowest classification score.

The classification of a test object $X_i = (x_1, x_2, ...x_d)$ containing d independent integers
15   is assigned to one of k gropus using the classification score defined as

$$D_k^2 = (X_i - \mu_k)^T M^{-1}(X_i - \mu_k)$$

where M^{-1} is the inverse of the pooled covariance matrix over all classes

$$M = \frac{1}{n} \sum_k \sum_{i=1}^{n_k} (X_i - \mu_k)(X_i - \mu_k)^T .$$

20   A third approach is backpropagating neural network analysis using a multilayer neural network with n input nodes, a hidden layer and an output layer. Neural network techniques have been widely reported and will not be discussed further here.

Other papers describe the use of elastic scattering spectroscopy in the diagnosis of a
25   number of conditions. Backman et al describe the detection of precancerous epithelial cells in "Detection of Preinvasive Cancer Cells", Nature, vol 406 p35 (2000). Perelman et al, in "Observation of Periodic Fine Structure in Reflectance from Biological Tissue: A New Technique for Measuring Nuclear Size Distribution", Phys. Rev. Lett. vol 80 p627 (1998) describe periodic fine structure in mucosal membranes.

The diagnosis of bladder cancer is described in "Spectroscopic Diagnosis of Bladder Cancer with Elastic Light Scattering" Mourant et al, Lasers in Surgery and Medicine, Volume 17 page 350 (1995). The use of elastic scattering to diagnose pathologies in the gastrointestinal tract is described in "Elastic Scattering Spectroscopy as a diagnostic tool for differentiating pathologies in the Gastrointestinal tract: preliminary testing", Mourant et al, Journal of Biomedical optics, Vol 1 p192, and in "Ultraviolet and visible spectroscopies for tissue diagnostics: fluorescence spectroscopy and elastic scattering spectroscopy", Bigio and Mourant Phys. Med. Biol. Volume 42 p803 (1997).

It is thus clear that the use of elastic scattering spectroscopy is attracting interest as a diagnostic tool. In spite of this research interest the most reliable approach presently used for detection of cancer in tissue and other conditions is histology. However, this is time consuming and laborious and in many situations, especially for diagnosing cancer, multiple biopsies may be needed.

There is thus a need to develop optical techniques further. One application is to guide conventional biopsies, avoiding false negatives and reducing the number taken while increasing yield. The long-term goal is to develop the techniques to a point where they can be used rapidly, efficiently and reliably to diagnose conditions without the need for histology.

According to the invention there is provided a method of processing a broadband elastic scattering spectrum obtained from tissue comprising the steps of: obtaining, in a plurality of fitting ranges of wavelength, fitting parameters giving the best fit to the spectrum in the respective fitting ranges; and recording the fitting parameters as a parameter data set representing the spectrum; wherein in at least one fitting range the fit is to the absorption of at least one predetermined component, and in the remainder of the fitting ranges the fit is to a smooth function.

By fitting in a number of different fitting regions to known absorption spectra and to a smooth function a measured spectrum including a large number of data points can be

reduced to the very much smaller number of data points, i.e. the fitting parameters. Subsequent data processing using the fitting parameters instead of the whole spectrum (as used in the prior art discussed above) may allow simpler, more reliable and more rapid assessment of the patient's condition.

5

The method can be thought of as using model dependent fitting, i.e. of analysing the spectrum using a model of the absorption with certain absorbing components absorbing at certain frequencies before carrying out any diagnosis or discrimination.

10   The fit to the absorption of at least one predetermined component may be to the absorption line shape of the at least one predetermined absorbing component. In particular, the fit may use a parabolic approximation to the peak of absorption of that absorbing component.

15   The fit to the absorption of at least one predetermined component may be a fit to an absorption spectrum previously measured using an optical biopsy probe on a sample of the predetermined absorbing component in a tissue-like matrix. This absorption spectrum, in general, differs from the simple absorption spectrum due to scattering obtained from a conventional optical transmission cell and available in most textbooks.

20   The use of a spectrum measured using an optical biopsy probe on a sample in a tissue-like substrate has not previously been suggested, as far as the inventor is aware.

Alternatively, especially for single component systems, the fit may be nothing more than determining the excess of absorption in the spectrum at a predetermined

25   frequency over the background spectral lineshape due to scattering calculated by extrapolating a straight line fit in a neighbouring region of the spectrum. The predetermined frequency is preferably the peak absorption wavelength of the absorbing component.

30   The fit to a smooth function is preferably to a straight line. Such fits are straightforward to carry out and with suitable choice of fitting ranges can parameterise

an absorption slowly varying with wavelength, i.e. the spectral lineshape due to scattering in the absence of any absorption features.

The fitting ranges, taken together, preferably include at least 60%, further preferably 80% of the wavelength range of the complete broadband elastic spectrum, at least in the range in which the spectrum has been measured with reasonable accuracy. In this way substantially all of the measured spectrum may be parameterised.

The method may also include, after obtaining fitting parameters in one fitting range, calculating a modified spectrum to compensate for the shape of the spectrum represented by the fitting parameters in that fitting range, and using the modified spectrum when fitting parameters in at least one further fitting range. This may be done by inputting the recorded fitting parameters in that fitting frequency range into a model of the absorption, calculating the expected absorption spectrum determined by the model with the input fitting parameters and subtracting the calculated absorption spectrum from the initial spectrum to obtain the modified spectrum used when fitting parameters in at least one further fitting range.

It should be noted that the fitting regions may overlap. For example, it may be desired to fit to a line shape of a known absorption component in a certain fitting region and then to multiply by a predetermined function to remove that line shape. However, there may still be information in the residual intensity of the lineshape due to scattering and this can be fitted using linear fitting parameters in a fitting region that may overlap or even be identical to the fitting region used to fit to the line shape of the absorption component.

The preprocessed spectrum may be fed to a discriminating algorithm to determine whether or not the data corresponds to one or more medical conditions. The discriminating algorithm may be trained to detect a particular medical condition or to discriminate between a number of clinically similar conditions. The training will use a number of training samples. The skilled person will realise that there are a number of suitable models with a number of variable fitting parameters for implementing the

discriminating algorithm. For example, a neural network approach may be used, a linear discriminant analysis or a hierarchical cluster analysis. All of these are known per se, and the first two are, for example, discussed in the paper by Zhenzhou et al mentioned above. It is generally the case, however, that the smaller number of model dependent fitting parameters obtained using the preprocessing method according to the invention can provide a benefit whatever fitting and diagnosis approach is used.

One reason for the improvement is the reduction in the number of points in the data set for fitting. Whether using a neural network or other discriminant analysis, the large number of points in the original data set of the whole spectrum means that a large number of training samples would be needed to train any model of the results output from the preprocessor. By reducing the number of points in the data set to be fitted less training is required and the fit can be carried out more reliably.

A preferred approach is hierarchical cluster analysis in which the n parameters of the spectrum define a unique point in n-dimensional space. Clusters of points are determined - the diagnosis corresponds to identifying in which cluster a measured spectrum point lies. Hierarchical cluster analysis has the advantage that it allows a "don't know" response, if for example the measured point is located far from any of the clusters identified. This is of advantage in preventing false diagnosis in cases where no such diagnosis can be reliably made.

A number of different absorbing components can be fitted and each absorbing component will absorb in a different wavelength range and hence a different fitting range.

One example of general application in human tissue is to fit to the haemoglobin absorption; this can be done by fitting to the saturation and the total hematocrit concentration. The saturation is defined as the percentage of oxygenated haemoglobin to the total hematocrit (oxygenated and deoxygenated haemoglobin). Such a fit to haemoglobin concentration may be carried out in a fitting range including at least part of the region of the spectrum from 320nm to 620nm.

Haemoglobin gives rise to an absorption feature which may dominate the spectrum in the region of 415nm, called the Soret band. Other constituents of tissue also absorb near this wavelength and once the "normal" Hb absorption has been removed the

5   absorption due to these features will be observed and may be fitted. The other features may be the absorption due to components such as cytokines.

Other absorbing components are relevant to a number of different kinds of spectrum. For example, to detect breast cancer it is preferred to fit to the beta-carotene absorption

10   spectrum in the fitting range 400-520nm.

In some studies, exogenous dye may be introduced into tissue for diagnostic purposes and accordingly the preprocessing method can include fitting to the spectrum of the dye used. For example for, for suspected cases of breast cancer, blue dye can be

15   introduced into human tissue to trace the spread of the disease. For the dye used in studies to date, Patent V blue dye (Trade Mark), a fitting range including at least past of the range 530nm to 720nm is suitable.

One predetermined fitting region may be a region in the range 630nm to 810nm, and

20   the fit may be a linear fit in this region. The method may also include fitting to a linear model in a number of other regions. These can include linear fitting in the range 340nm to 360nm, and/or the range 320nm to 330nm. These fittable regions will be observed after the removal of absorption features.

25   The spectral trace may be checked for and the spectrum rejected if the check reveals measurement errors or unsuitable data. For example, it may be advisable to check for a minimal transmitted intensity in the Soret band and to reject the spectrum if the measured transmitted intensity is substantially zero in this band. In other words, if the measured transmitted intensity is less than 10% full scale, preferably less than 5% full

30   scale, the spectrum may be rejected. Another possibility is to check for interference from background illumination and to reject the spectrum if background illumination is too high. Further, the spectrum can be checked for contact between probe and tissue.

The invention also relates to a method including recording an elastic scattering spectrum from tissue, and preprocessing the spectrum as described above.

5   The tissue may be in vivo, i.e. tissue incorporated in the living human body.

The tissue may be in vitro, i.e. tissue removed from the body.

In another aspect there is provided a method comprising the steps of recording an
10   elastic scattering spectrum; preprocessing the spectrum using a preprocessing method to obtain a plurality of fitting parameters characterising the spectrum; testing the preprocessed spectrum using a discriminant model; and outputting a result based on the model.

15   The result may be an output indicating to which class, if any, the recorded elastic scattering spectrum belongs. In embodiments, the output may be a diagnosis.

In embodiments, the discriminant model may be a neural network, linear discrimination, hierarchical cluster analysis or other methods as are known to those
20   skilled in the art.

The preferred discriminant model uses hierarchical cluster analysis. This groups data into unbounded class regions permitting "not sure" diagnostic indications, rather than forcing a decision and risking a false diagnosis.

25

In another aspect, the invention relates to a method comprising the steps of recording an elastic scattering spectrum from tissue; processing the spectrum to produce a number of parameters characterising the spectrum; determining to which, if any, of a number of classes the parameterised spectrum belongs; and outputting the class, if any,
30   to which the spectrum is determined to belong.

In a further aspect there is provided a training method comprising the steps of recording a plurality of optical biopsy spectra from tissue for which it is known whether the tissue displays a predetermined medical condition; preprocessing each of the spectra using a method as defined above; and training a discriminant model using

5   the preprocessed spectra.

In another aspect there is provided an apparatus for optical biopsy, comprising apparatus for elastic scattering spectroscopy of tissue, including a light source for emitting light over a broad range of frequencies; a probe for transmitting light from the

10   light source to tissue and for receiving light scattered in the tissue; a spectrometer for measuring the intensities of the received light as a function of frequency; and a processor for processing the measured light spectrum arranged to carry out the method as described above.

15   The apparatus may include a first optical fiber bringing light from the light source to a probe tip; and a second optical fiber bringing scattered light from the probe tip to the spectrometer; wherein the ends of the first and second fibers at the probe tip are arranged adjacently spaced apart by a predetermined distance.

20   The apparatus may include a decision processor for checking the fitted parameters against the results for one or more predetermined medical conditions and outputting the best fit medical condition based on the decision processor output.

Specific embodiments of the invention will now be described, purely by way of

25   example, with reference to the accompanying drawings in which:

Figure 1 shows an optical biopsy system;

Figure 2 is a flow diagram of a preprocessing method according to a first embodiment of the invention;

Figure 3 is a flow diagram of an optical biopsy method using the preprocessing of

30   Figure 2;

Figure 4 is a first example spectrum, taken from normal tissue;

Figure 5 is an example spectrum showing interference from background illumination;

Figure 6 is an example spectrum showing a small peak around 690nm;

Figure 7 illustrates normalisation;

Figure 8 shows haemoglobin absorption; and

Figure 9 shows the spectrum after haemoglobin absorption has been compensated.

The first step is to record an elastic scattering spectrum. Referring to Figure 1, the apparatus for recording the spectrum includes an excitation light source 1 and a probe 3. A spectrometer 5 for splitting light and a detector array 7 for measuring the intensity of the split light are also provided. A first optical fibre 9 transmits light from the light source 1 to the end 11 of the probe and a second optical fibre 13 picks up light from the end of the probe and transmits it to the spectrometer for measurement. A computer 15 including interface electronics is electrically connected to the light source 1 the spectrometer 5 and the detector 7 for controlling these parts. The apparatus may be as described in US5,305,026 (discussed above) but this is not required and the skilled person will readily conceive of alternative probes, light sources and spectrometer arrangements.

In use, the end 11 of the probe is brought up to a tissue sample, such as the skin of a patient, so that the ends of the first and second optical fibre are adjacent to tissue. Light is then emitted from the excitation light source, passes through the first optical fiber 9 and into the tissue. After passing through and being scattered in the tissue some of the scattered light enters the second optical fiber 13 and passes to the spectrometer where the spectrum is measured.

In the specific example the excitation light source 1 is a xenon arc light which emits a number of pulses down the send fiber. The output detected at the spectrometer 5 is integrated to catch the response from the plurality of pulses. The spectrum is "auto-ranged" to ensure that the peak intensity at a chosen wavelength is scaled to lie above some threshold wavelength but below the saturation level of the detector 7, in this instance a CCD, by varying the number of light pulses. A second spectrum without illumination is obtained immediately before or after the illuminated spectrum is taken and the second spectrum is subtracted from the illuminated spectrum This removes

effects due to extraneous light sources such as room lights, operating theatre lights or the headlamps of an endoscope.

The measured spectrum is then ratioed to a reference spectrum which is that obtained from a white material with constant reflection properties from the UV to the IR. Such a material is commercially available under the trade name "Spectralon". The resultant spectrum is then a ratio to a nominal mean intensity level 100. This removes any effects due to the spectrum of the light source. Next, the spectrum is smoothed using a simple "boxcar" function with a unit length of 7 pixels.

An example of the spectrum thus achieved is shown in Figure 4. This is the spectrum that forms the starting point for preprocessing. The processing of the spectrum is carried out by a program 17 stored in the computer 15 for causing the computer to carry out the steps of the method that will now be described which will now be described with reference to Figure 2.

The first step 102 is to define the useable wavelength range. This is principally determined by the noise in the spectrum and generally lies in the range 320-810nm. Results outside this predetermined window are rejected. The skilled person will realise that when using different light sources to the Xenon light or different probe components useable results may be achieved over a different wavelength range and in that case a different wavelength window can be used.

The spectrum is then checked (Step 104) for measurement errors. In particular, it is common for the absorption peaks of oxygenated haemoglobin HbO_2 and deoxygenated haemoglobin Hb at 415nm and 433nm to be saturated if too much blood is present in the tissue sample. That is to say, the measured light intensity may be substantially zero in this region. If this is the case the absorption is judged to be too great and the spectrum rejected.

The next check, in step 106, is to determine if the signal has been corrupted by background illumination. This can occur in one of two ways. If the illumination is too

12

bright the CCD may become saturated. This can easily be checked for by determining if the peak signal level is too low; if it is the spectrum is rejected.

Another possibility is that sample movement between the tissue sample and the probe has occurred between the taking the illuminated and the dark measurements. This is a particular problem for endoscopic measurements made in the gut since gut tissue is highly active. This effect has been observed to give rise to a peak or trough between 600 and 655nm. The existence of such a peak of trough is detected and if present the sample is rejected. Alternatively the recorded spectrum may be modified to eliminate this artefact.

A third check is then carried out in step 108 for good contact between probe and tissue. Imperfect contact appears to give a V-shaped feature at 690nm, as illustrated in both spectra shown in Figure 6, which can be checked for. Again, if the feature is present the spectrum can be rejected.

Assuming that the spectrum is acceptable after the above checks the next step in this example is to fit linear regions. It appears that the spectrum is approximately linear between 630nm and 810nm but better results are obtained by fitting in two regions, from 740 to 810nm (step 110) and from 630 to 710nm (step 114). There are a number of reasons for this. As mentioned below, blue dye can disturb the region below 730nm. Secondly, the haemoglobin tail can extend to well above 630nm. Thirdly, absorption due to water gives rise to a small non-linearity around 730nm.

The fitting is carried out using conventional linear regression to fit a region of the spectrum to a straight line gradient m and intercept b. As is known, m and b are given by:

$$m = \frac{n\sum_{i}^{n}x_{i}y_{i}\sum_{i}^{n}x_{i}\sum_{i}^{n}y_{i}}{n\sum_{i}^{n}x_{i}^{2}-\left(\sum_{i}^{n}x_{i}\right)^{2}}$$

$$b = \frac{\sum_{i}^{n}y_{i}\sum_{i}^{n}x_{i}^{2}-\sum_{i}^{n}x_{i}\sum_{i}^{n}x_{i}y_{i}}{n\sum_{i}^{n}x_{i}^{2}-\left(\sum_{i}^{n}x_{i}\right)^{2}}$$

where the subscripts i represent summation over the points in the fitted region.

Between fitting these two steps any absorption caused by blue dye may be removed in step 112. When used for the identification of sentinel nodes in breast tissue a blue dye, known as Patente Bleu V (Trade Mark) is used which has an absorption band with a peak at 635nm. The blue dye spectrum is removed in the same way as used for fitting to haemoglobin features, as explained below.

After fitting to the 630nm to 710 nm region the spectrum is normalised (step 116) to the gradient in the 630-710nm linear section. In other words, the ratio to a line of constant gradient is taken for the whole spectrum so that the region of the spectrum between 630 and 710nm becomes flat. This is illustrated in Figure 7 which shows the original spectrum, a line of constant gradient fitted in the region between 630nm and 710nm and the normalised spectrum normalised so that the graph is flat in the region 630nm to 710nm..

Next, in step 118, haemoglobin lines are fitted and compensated.

In general, optical absorption is given by the well-known Beer-Lambert law, which states that the intensity of light remaining in a beam that has passed through a thickness z of material which has an absorption coefficient of γ is given by $I=I_0e^{-z\gamma}$,

14

where I_0 is the incident intensity. Conventional absorption spectra can readily be obtained in the published literature.

However, in tissue, the situation is more complicated and scattering effects must be taken into account. In elastic scattering spectroscopy, the situation is more complex still and the path-length through which photons travel through tissue is not well defined. Indeed, the path length varies non-linearly with the wavelength of light. Accordingly, published spectra of components may not be suitable and the absorption spectra used in the present invention are spectra measured using the equipment described above and in-vitro tissue phantoms. The spectra used thus implicitly compensate for the geometry and so the accuracy of absorption measurement and removal is enhanced.

For simple, one-component absorption, the amount of a component may be estimated using an offset from a straight line fit to an adjacent region. In the embodiment described, this is done by normalising the graph in the adjacent region and then calculating the offset from this absorption at the known frequency peak. This process is used for blue dye and beta carotene in the described embodiment, but as the skilled person will realise the process can also be used for other one-component systems.

For haemoglobin there is more than one absorber (Hb and HbO_2) and the amount of haemoglobin must be measured using some form of fit. This is done by fitting the peak regions to a parabolic line shape and using a regression analysis in a known way.

Accordingly, the peaks in the absorption spectrum at which the measurements should be taken are defined. Referring to Figure 8, which shows absorption spectra for oxygenated (HbO_2) and deoxygenated haemoglobin (Hb), differences in the absorption profiles are obvious. After the normalisation of step 116, the wavelength positions and absorption coefficients in the vicinity of peaks in this spectrum are fitted to using a local parabolic approximation and multiple linear regression analysis.

It may be useful to find points of equal absorption for both components and the absorption values for the second spectrum at positions corresponding to peaks in the first.

5   Once these values have been found as fit parameters these can be converted into absorption coefficients. The fitting values be used to calculate the saturation, defined as the percentage of oxygenated haemoglobin to total haemoglobin concentration, known as hematocrit.

10   The fitting parameters are then used to compensate the absorption represented by these fitting parameters and subtract that from the spectrum. The absorption calculated in the model from the fitted concentrations of haemoglobin is determined and subtracted (step 120) from the measured spectrum to obtain the spectrum used in subsequent steps. Two pairs of spectra before and after this subtraction are shown in figure 9.

15   Spectrum 91 (before) becomes spectrum 93 (after), and spectrum 95 (before) becomes spectrum 97 (after) illustrated in figure 9.

The process of removal used in the embodiment relies on the assumption that the Beer-Lambert function describes the absorption sufficiently well and that the in-vitro

20   absorption spectrum does not differ substantially to the in-vivo spectrum. The removal process is straightforward. The model used is simply to reverse the model used in the initial absorption process. Each point in the model spectrum is simply multiplied by a factor obtained from the fitted concentration by inverting Beer-Lambert's law to determine a model spectrum that can then be subtracted from the

25   measured spectrum.

Although haemoglobin shows marked non-linearities in absorption at high concentrations this should not present a problem in samples that pass the test of step 104 above and have sufficiently little haemoglobin.

30
Although the subtraction does not, in general, completely eliminate the haemoglobin peaks, and indeed the effect is sometimes to invert the peaks, what the step does do is

to remove from the spectrum those features already recorded by the haemoglobin fitting parameters.

The skilled person will realise that there are a number of alternative ways to calculate
5   the haemoglobin concentrations from the measured data. Rather than just fitting to the peaks, an alternative is to calculate the saturation from the isobectic point (the point at which both components are absorbed equally. The absolute absorption might then be used to calculate the total hematocrit.

10 In step 122, a linear fit is made in the region 540 to 630nm.

In step 124, the absorption due to beta-carotene is fitted, recorded and then the absorption spectrum of beta-carotene subtracted from the measured spectrum. This fit takes place in a fitting range beginning at around 520nm and extending down to
15 include at least distinct peaks at 480nm, 450nm and 420nm. Accordingly, the fitting range used may be all or part of the range 400nm to 520nm. Fitting to a Beta-carotene peak and consequently obtaining fitting parameters related to the concentration of Beta carotene is particularly important in pre-processing data from breast tissue, since beta carotene is related to vitamin C and found only in fat. Its presence is a contra-
20 indication to malignancy.

Linear regions between 490nm and 520nm and between 455nm and 480nm are fitted. (step 126). As before, the gradients, intercept intensities and regression statistics are recorded.
25

Then in step 128, the residual absorption to other chromophoric components within the region of the Hb Soret band at around 415nm is measured by parabolic approximation. This gives an indication of the different profile of cytokines and other tissue absorbers in the tissue as apposed to normal, whole blood.
30

Next, linear fits are provided in the regions of 340 to 360nm and also 320 to 330nm.

17

Although it may appear that a large number of fits have been carried out, the number of data points required to parameterise all of the fits is very much less than the number originally measured, which in the apparatus used is of order 1800 data points. The reduced number of data points can readily be used as the inputs to subsequent models, for example to train a model or to diagnose based on a previously trained model.

In order to use the data measured for diagnosis, it is first necessary to "train" a model by providing it with a number of fitted spectra for which it is known whether the tissue is normal or has a given medical condition in order to then be able to then use the test to determine whether another patient has that condition without the need for histology.

In the preferred embodiment, hierarchical cluster analysis is used. In hierarchical cluster analysis points in an n-dimensional space are grouped into a hierarchy of clusters. The grouping may occur from the bottom up, in which case each point is assigned to a single point cluster, and an algorithm groups pairs of clusters one after the other to produce a family tree of clusters. Alternatively, it is also known to successively divide clusters to produce a hierarchy from the top down.

Hierarchical cluster analysis is known, for example for face recognition and other pattern recognition. Its use in diagnosis from elastic scattering spectra is mentioned in Paul M Ripley, D. Pickard et al. "A Comparison of Artificial Intelligence Techniques for Spectral Classification in the Diagnosis of Human Pathologies based upon Optical Biopsy" Novel Biomedical Optical Spectroscopy, Imaging and Diagnostics (Optical Society of America, Apr 200, Miami, FL) OSA Biomedical Topical Meetings Proceedings, 2000, MC5.

To train the model, a number of test samples of known diagnosis are taken (step 140); the samples are known as the training set. The training samples are then preprocessed as described above (step 142), and divided into clusters using hierarchical cluster analysis (step 144). Hierarchical cluster analysis is described, for example, in Duda, Hart and Stork "Pattern classification: 2nd edition", John Wiley & Sons, California,

18

USA, 1998, and Andre Hardy "On the number of clusters", Computational Statistics & Data Analysis, volume 23 pages 83-96, 1996.

Each of the clusters is assigned a rating based on the points in the cluster. For example, if most of the points in the cluster based on the training set have the given medical condition, the cluster is labelled "suspicious". Conversely, if the majority of points in the cluster are clear of that condition, the cluster is labelled "clear".

Then, when an elastic scattering spectrum of a patient is taken, the point represented by the fitting parameters is determined and its Euclidean distance to each cluster is determined in the multidimensional parameter space spanned by the fitting parameters. If the point is further than a predetermined distance from any cluster, then a "don't know" result is given. Otherwise, the point is close to a cluster and it is assigned to the closest cluster and the label corresponding to that cluster is output as a diagnosis. This technique is known as the "leave one out" approach in hierarchical cluster analysis.

The method of the invention is particularly useful in diagnosing a number of conditions, for example breast cancer, dysplasia within the gastro-intestinal tract, cancers of the oral mucous, cervical cancer, lung cancer and skin cancer.

Although the invention has been described with reference to a specific example, the skilled person will realise that a number of variations are possible. In particular, a number of other predetermined components may be fitted for.

For example, a fit to the absorption line shape of bile pigment may be carried out in a fitting range including at least part of the range from 230nm to512nm .

A fit to melanin absorption may be carried out using a large fraction of the whole spectrum as the fitting range; the fit may be carried out after fits to other absorbing components have been carried out and the absorption due to the other components subtracted from the measured spectrum.

A fit to protein absorption may be carried out below 350nm.

Furthermore, although the invention as described above uses the wavelength as the abscissa of the spectrum, it is possible instead to use frequency or any other parameter related to wavelength as the abscissa.

CLAIMS

1.   A method of processing a broadband elastic scattering spectrum comprising the steps of:

obtaining, in a plurality of fitting ranges of wavelength, fitting parameters giving the best fit to the spectrum in the respective fitting ranges; and

recording the fitting parameters as a parameter data set representing the spectrum;

wherein

in at least one fitting range, but not all fitting ranges, the fit is to the absorption of at least one predetermined component; and

in the remainder of the fitting ranges, the fit is to a smooth function.

2.   A method according to claim 1 wherein the fit to a smooth function is to a straight line.

3.   A method according to claim 1 or 2 wherein the plurality of fitting ranges, taken together, include at least 60% of the wavelength range of the complete broadband elastic scattering spectrum.

4.   A method according to any preceding claim wherein in at least one fitting range, the fit to the absorption of at least one predetermined component is to the absorption line shape of the at least one predetermined absorbing component.

5.   A method according to claim 4 wherein the fit to the absorption line shape of the at least one predetermined absorbing component uses a parabolic approximation.

6.   A method according to any preceding claim wherein the fit to the absorption of at least one predetermined component is a fit to an absorption spectrum previously measured using an optical biopsy probe on a sample of the predetermined absorbing component in a tissue-like matrix.

7.   A method according to any preceding claim further comprising, after the step of obtaining fitting parameters in one fitting range, calculating a modified spectrum to compensate for the shape of the spectrum represented by the fitting parameters in that fitting range, and using the modified spectrum when fitting parameters in at least one further fitting range.

8.   A method according to claim 7 wherein after obtaining fitting parameters representing a fit to an absorption line shape in one fitting range, a modified spectrum is calculated by inputting the recorded fitting parameters in that fitting frequency range into a model of the absorption, calculating the expected absorption spectrum determined by the model with the input fitting parameters and subtracting the calculated absorption spectrum from the initial spectrum to obtain the modified spectrum used when fitting parameters in at least one further fitting range.

9.   A method according to any preceding claim wherein, in at least one fitting range, the predetermined absorbing components are haemoglobin in oxygenated and deoxygenated forms.

10.   A method according to claim 9 wherein the fitting range in which the predetermined absorption components are oxygenated and deoxygenated haemoglobin is at least part of the range from 320nm to 620nm.

11.   A method according to any preceding claim wherein in one fitting range in the range 400 to 520nm the predetermined absorbing component is beta-carotene.

12.   A method according to any preceding claim including calculating absorption of other chromophoric components in the tissue which share haemoglobin's Soret absorption band around 412 and 430nm and recording the differential absorptions as fitting parameters.

13.   A method according to claim 12 wherein the chromophoric components include cytokines.

14.   A method according to any preceding claim wherein in a fitting range including at least part of the range from 230nm to 512nm the predetermined absorbing component is a bile pigment.

15.   A method according to any preceding claim wherein in one fitting range covering substantially the whole of the spectrum the predetermined absorbing component is melanin.

16.   A method according to any preceding claim including fitting in a range below 350nm to protein absorption.

17.   A method according to any preceding claim wherein in one fitting range in the predetermined absorbing component is an exogenous dye.

18.   A method including recording an elastic scattering spectrum from tissue, and processing the spectrum using a method in accordance with any preceding claim.

19.   A method according to any preceding claim further comprising:
        determining to which, if any, of a number of classes the parameterised spectrum belongs by measuring the distance in a multidimensional parameter space between the parameters characterising the spectrum and parameters characterising the classes.

20.   A method according to claim 19 wherein the step of determining to which, if any class the parameterised spectrum belongs uses classes determined by hierarchical cluster analysis.

21.   A method comprising the steps of:
        recording an elastic scattering spectrum from tissue;
        processing the spectrum to produce a number of parameters characterising the spectrum;

determining to which, if any, of a number of classes the parameterised spectrum belongs by measuring the distance in a multidimensional parameter space between the parameters characterising the spectrum and parameters characterising the classes; and

5   outputting the class, if any, to which the spectrum is determined to belong,

wherein the step of determining to which, if any class the parameterised spectrum belongs uses classes determined by hierarchical cluster analysis.

22.   A computer program product for loading into a computer arranged to cause the

10   computer to carry out the steps of a method according to any preceding claim.

23.   A computer program product recorded on a data carrier for carrying out the steps of a method according to any of claims 1 to 21.

15   24.   Apparatus for elastic scattering spectroscopy of tissue, including

a light source for emitting light in a broad frequency band;

a probe for transmitting light from the light source to tissue and for receiving light scattered in the tissue;

a spectrometer for measuring the spectrum of the received light as a function of

20   frequency; and

a processor for processing the measured light spectrum arranged to carry out the method of any of claims 1 to 17 to obtain a plurality of fitted parameters characterising the spectrum.

25   25.   Apparatus according to claim 24 wherein the probe contains a first optical fiber bringing light from the light source to a probe tip; and

a second optical fiber bringing scattered light from the probe tip to the spectrometer; wherein the ends of the first and second fibers at the probe tip are arranged adjacently spaced apart by a predetermined distance.

30
26.   Apparatus for analysing an elastic scattering spectrum taken from tissue, comprising

24

a data store for recording the elastic scattering spectrum;

a processor arranged to carry out the steps of any of claims 1 to 21 on the spectrum stored in the data store to obtain a plurality of fitted parameters characterising the spectrum.

27.   Apparatus according to any of claims 24 to 26 further comprising a decision processor for checking the fitted parameters against the results for one or more predetermined states and outputting the best fit state based on the decision processor output.

28.   Apparatus according to claim 27 wherein the decision processor uses hierarchical cluster analysis.

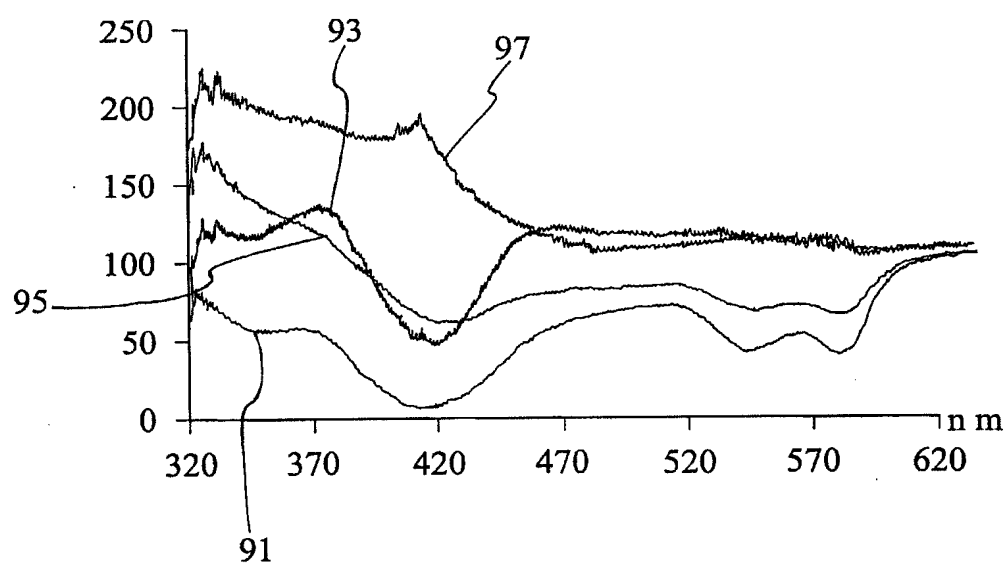29.   Apparatus according to claim 27 or 28 wherein the states are medical conditions.

30.   A training method comprising the steps of

recording a plurality of broadband elastic scattering spectra from patients for which it is known whether they have a predetermined medical condition;

preprocessing each of the spectra using a method as defined in any of claims 1 to 17; and

training a discriminant model using the preprocessed spectra.

FIG 1

Record Spectra

Preprocess Each Spectrum

Train Discriminant
Model

FIG 3
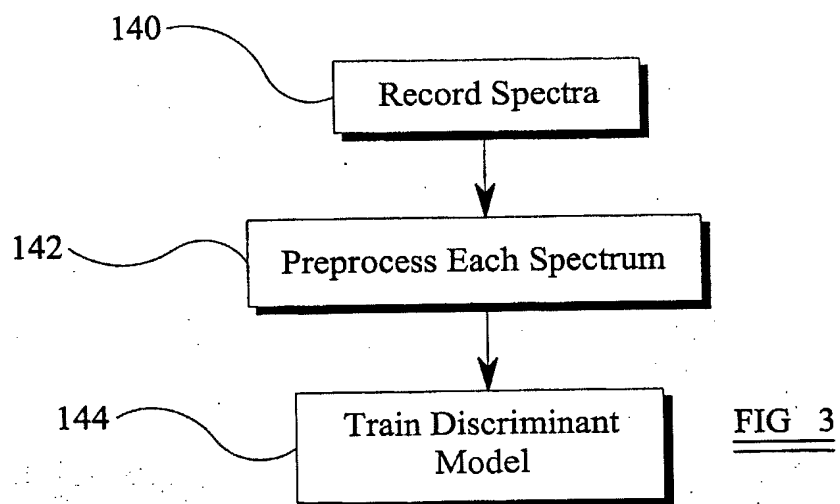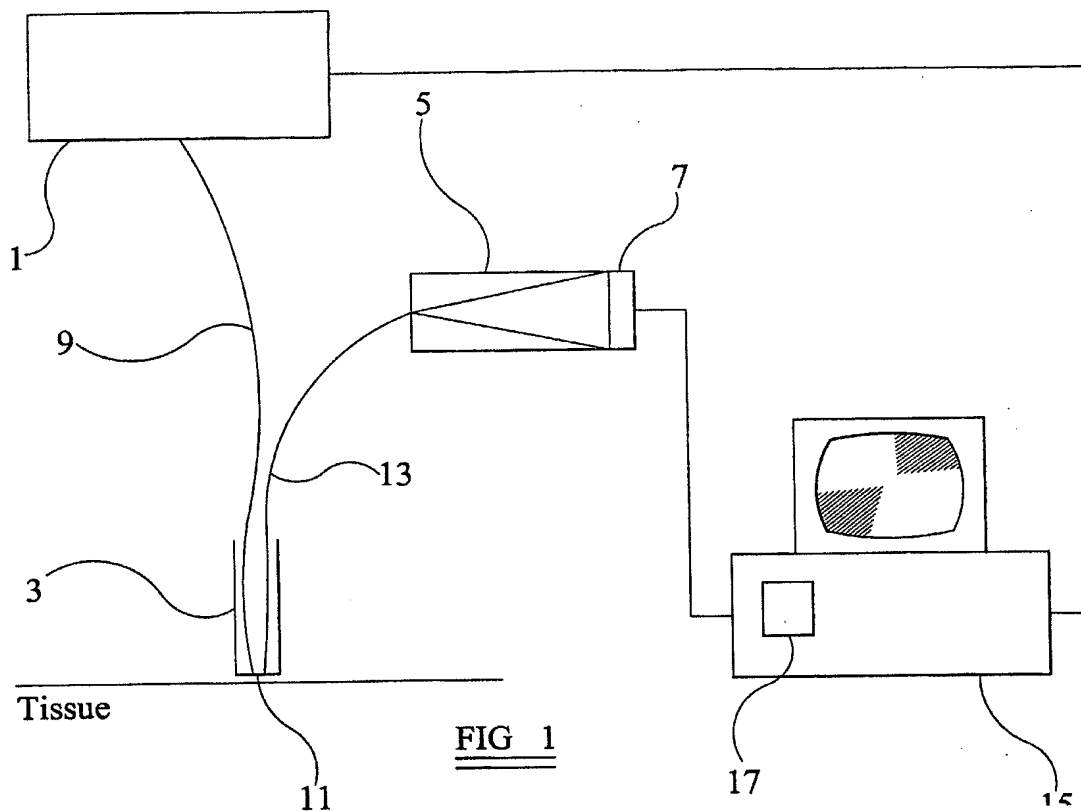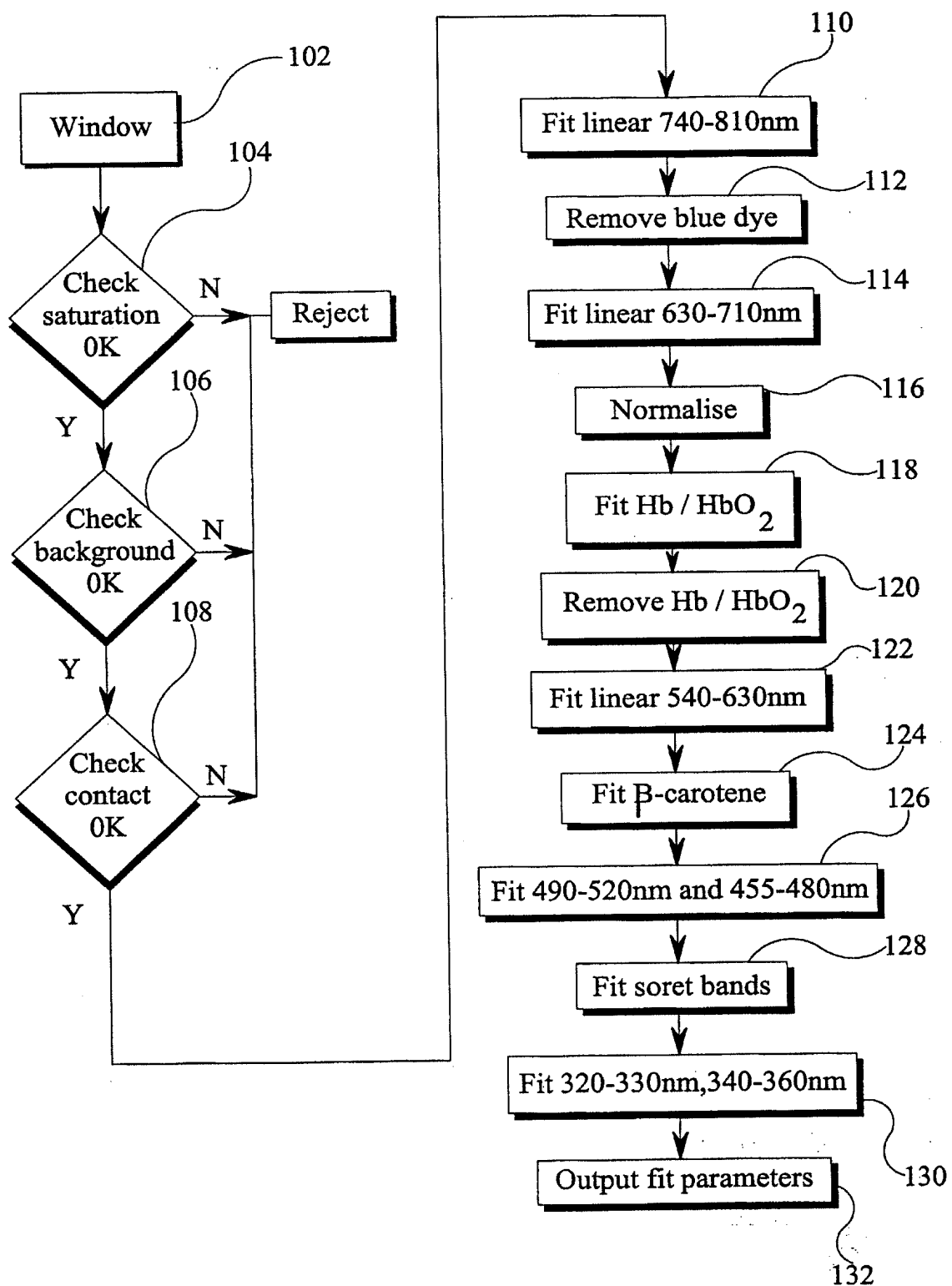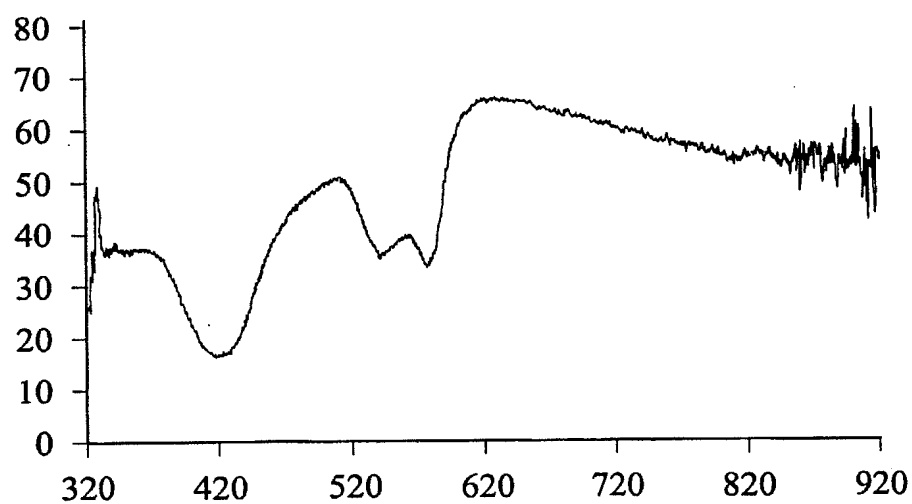
FIG 4
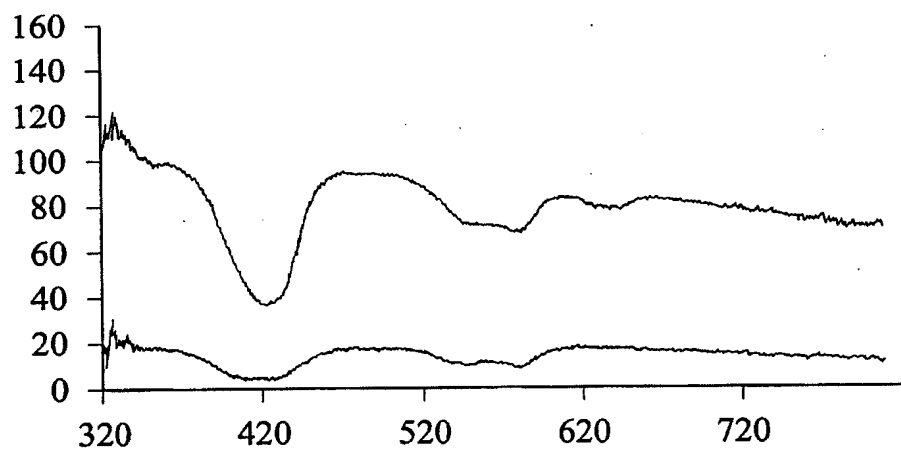
FIG 5
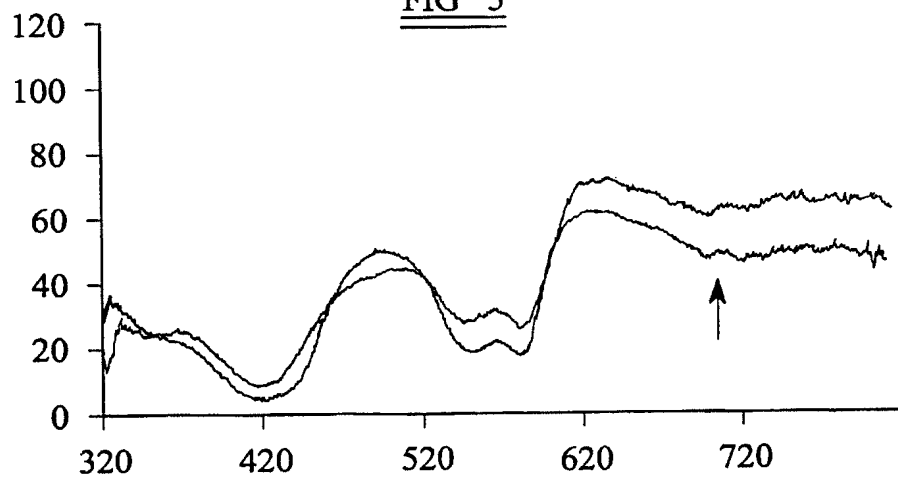
FIG 6
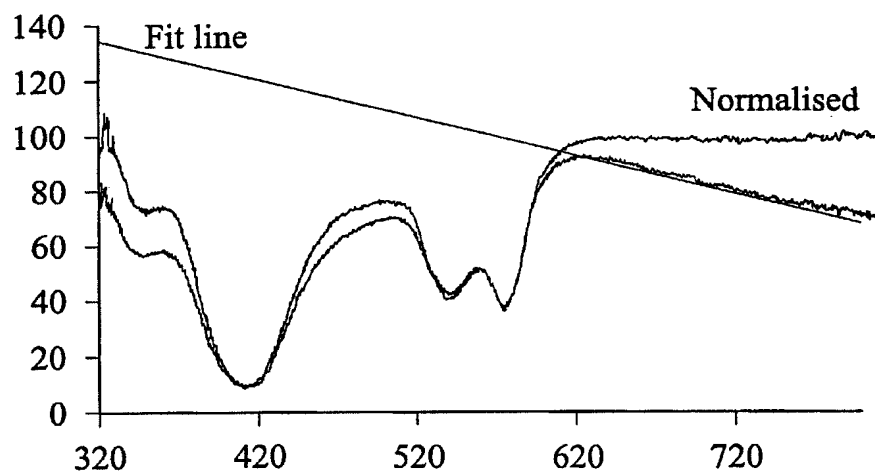
FIG 7
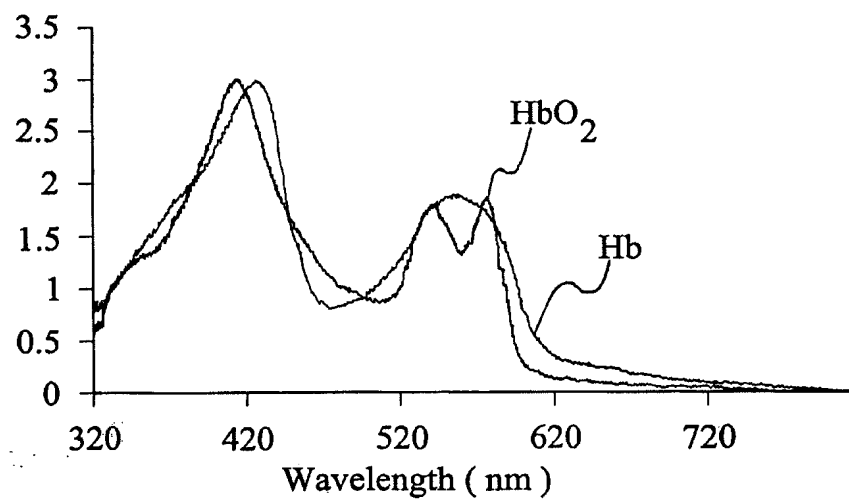
FIG 8

FIG 9

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7   G01N21/49   A61B5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7   G01N   A61B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, INSPEC, COMPENDEX, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category°	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CASTANHO M A R B ET AL.: "Separating the turbidity spectra of vesicles from the absorption spectra of membrane probes and other chromophores" EUROPEAN BIOPHYSICS JOURNAL, vol. 26, 1997, pages 253-259, XP002196937	1-4,7, 18,22,23
Y	the whole document	6,9-17, 19,20, 24,25, 27-30
A	────	5
Y	EP 0 906 936 A (STICHTING TECH WETENSCHAPP) 7 April 1999 (1999-04-07) abstract	6
	────	
	-/--	

[X] Further documents are listed in the continuation of box C.   [X] Patent family members are listed in annex.

° Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 April 2002	15/05/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hoogen, R

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	MOURANT J R ET AL.: "Elastic scattering spectroscopy as a diagnostic tool for differentiating pathologies in the gastrointestinal tract: preliminary testing" JOURNAL OF BIOMEDICAL OPTICS, vol. 1, no. 2, April 1996 (1996-04), pages 192-199, XP001069378 cited in the application	21,26
Y	* sections 1, 2, and 3.1	19,20, 24,25, 27-30
Y	BIGIO I J ET AL.: "Diagnosis of breast cancer using elastic-scattering spectroscopy: preliminary clinical results" JOURNAL OF BIOMEDICAL OPTICS, vol. 5, no. 2, April 2000 (2000-04), pages 221-228, XP001070060 the whole document	9-17

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
EP 0906936	A	07-04-1999	NL	1006902 C2	02-03-1999
			EP	0906936 A1	07-04-1999
			US	6224969 B1	01-05-2001